

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2002, 11:48:41 ; Search time 50.08 Seconds  
(without alignments)  
37.970 Million cell updates/sec

Title: US-09-432-546-4

Perfect score: 99

Sequence: 1 RRMWMPWKWPLI 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

1: SP:archaea:*
2: SP:bacteria:*
3: SP:fungi:*
4: SP:human:*
5: SP:invertebrate:*
6: SP:mammal:*
7: SP:mhc:*
8: SP:organelle:*
9: SP:phage:*
10: SP:plant:*
11: SP:rodent:*
12: SP:virus:*
13: SP:vertebrate:*
14: SP:unclassified:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57.6	57.6	1662	2 P71431	P71431 leptochoix
2	56.6	56.6	236	2 Q9KFN3	Q9KFN3 bacillus ha
3	55.5	56.1	165	10 Q9SNM3	Q9SNM3 oryza sativ
4	55.5	56.1	676	10 Q9FGZ6	Q9FGZ6 arabidopsis
5	53.5	53.5	95	10 Q9LQNO	Q9LQNO arabidopsis
6	52.5	53.0	970	10 Q80790	Q80790 arabidopsis
7	52.5	53.0	1231	10 Q82276	Q82276 arabidopsis
8	52.5	53.0	157	5 Q9Y0E8	Q9Y0E8 drosophila
9	52.5	53.0	301	2 Q915M4	Q915M4 pseudomonas
10	52.5	52.5	1245	3 Q9Y7V5	Q9Y7V5 trichoderma
11	51.5	52.0	109	10 Q40548	Q40548 nicotiana t
12	51.5	52.0	196	10 Q08195	Q08195 nicotiana t
13	51.5	52.0	209	10 Q08194	Q08194 nicotiana t
14	51.5	52.0	321	10 Q9S7Z8	Q9S7Z8 arabidopsis
15	51.5	52.0	2970	12 Q56073	Q56073 hepatitis g
16	51.5	51.5	1173	12 Q990M4	Q990M4 human coron
17	51.5	51.5	1173	12 Q990M3	Q990M3 human coron
18	51.5	51.5	1173	12 Q990M2	Q990M2 human coron
19	51.5	51.5	1173	12 Q990M1	Q990M1 human coron

20	51	51.5	1383	12 Q84712	Q84712 porcine epi
21	50.5	51.0	560	5 Q44626	Q44626 caenorhabdi
22	50	50.5	83	2 Q9WYF1	Q9WYF1 thermotoga
23	49.5	50.0	257	2 Q56924	Q56924 yeastina en
24	49.5	50.0	491	4 Q9B012	Q9B012 homo sapien
25	49.5	50.0	723	12 Q9D0C4	Q9D0C4 tt virus. o
26	49	49.5	141	11 Q9CZAI	Q9CZAI mus musculu
27	49	49.5	169	11 Q9Y9Q4	Q9Y9Q4 aeryopyrum p
28	49	49.5	251	2 Q52853	Q52853 bacillus su
29	49	49.5	467	5 Q19573	Q19573 caenorhabdi
30	49	49.5	492	10 Q9AV15	Q9AV15 oryza sativ
31	49	49.5	669	5 Q18300	Q18300 caenorhabdi
32	49	49.5	746	12 Q9JH31	Q9JH31 tt virus. o
33	49	49.5	1053	4 Q9BXX0	Q9BXX0 homo sapien
34	48	48.5	196	10 Q9LP00	Q9LP00 arabidopsis
35	48	48.5	236	4 Q9P0T8	Q9P0T8 homo sapien
36	48	48.5	314	10 Q9Z0Y7	Q9Z0Y7 arabidopsis
37	48	48.5	485	2 Q9AFK1	Q9AFK1 shigella fl
38	48	48.5	947	10 Q9LDL8	Q9LDL8 arabidopsis
39	48	48.5	1172	10 Q9LP05	Q9LP05 arabidopsis
40	47.5	48.0	111	5 Q18753	Q18753 caenorhabdi
41	47.5	48.0	752	2 Q9K0M8	Q9K0M8 vibrio chol
42	47	47.5	73	2 Q9RW06	Q9RW06 delnoccocus
43	47	47.5	133	5 Q9VX67	Q9VX67 drosophila
44	47	47.5	138	11 Q9D3H6	Q9D3H6 mus musculu
45	47	47.5	233	10 Q9S181	Q9S181 arabidopsis

#### ALIGNMENTS

```

RESULT 1
P71431 PRELIMINARY; PRT: 1662 AA.
ID P71431
AC P71431;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DR MORA GENE ENCODING MANGANESE OXIDIZING PROTEIN PRECURSOR.
GN MORA.
OS Leptochrix discophora.
OC Bacteria; Proteobacteria; beta subdivision; Comamonadaceae;
OC Leptochrix.
OX NCBI_Taxid=89;
RN [1]
RP SEQUENCE OF 1-1150 FROM N.A.
RC STRAIN=SS-1;
RA Corstjens P.L.;
RL Thesis (1993), Biochemistry, Leiden University, The Netherlands.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SS-1;
RA Corstjens P.L.;
RL Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=SS-1;
RA Corstjens P.L.A.M., de Vrind J.P.M., Goosen T., de Vrind-de Jong E.W.;
RL Geomicrobiol. J. 14:91-108(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=SS-1;
RA Corstjens P.L.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z25774; CA81037.1; -.
DR InterPro: IPR001865; Ribosomal_S2.
DR PROSITE: PS00962; RIBOSOMAL_S2_1; UNKNOWN_1.
KW SIGNAL.
FT SIGNAL. 1 33 POTENTIAL.
FT CHAIN. 34 1662
SQ SEQUENCE 1662 AA; 174292 MW; 639E23860DD9246 CRC64;

```

Query Match 57.6%; Score 57; DB 2; Length 1662;  
 Best Local Similarity 63.6%; Pred. No. 16;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 RMPWMPKMP 12  
 |||||  
 Db 490 RMHWGPMFMPV 500

## RESULT 2

ID 09KFN3 PRELIMINARY; PRT; 236 AA.  
 AC 09KFN3;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE BH0446 PROTEIN.  
 GN BH0446.  
 OS Bacillus halodurans.  
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 CC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kihara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 DR EMBL: AP001508; BAB04165.1; -  
 KM Complete proteome.  
 SQ SEQUENCE 236 AA; 26988 MW; A0BD60D9FE34FA9 CRC64;

Query Match 56.6%; Score 56; DB 2; Length 236;  
 Best Local Similarity 77.8%; Pred. No. 3.7;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 WMPKMP 13  
 |||||  
 Db 192 WLPKMP 200

## RESULT 3

ID 09SNN3 PRELIMINARY; PRT; 165 AA.  
 AC 09SNN3;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE HYPOTHETICAL PROTEIN.  
 OS Oryza sativa (Rice).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 CC Eriocaridaceae; Oryzaeae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. NIPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC  
 RT clone:p0493C11.";  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP000559; BAA84796.1; -  
 SQ SEQUENCE 165 AA; 18358 MW; 9EE82A558ED22D95 CRC64;

Query Match 56.1%; Score 55.5; DB 10; Length 165;  
 Best Local Similarity 77.8%; Pred. No. 3.1;  
 Matches 7; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 4 PMWMP-KMP 11  
 |||||  
 Db 40 PMWMP-CMP 48

## RESULT 4

ID 09FG26 PRELIMINARY; PRT; 676 AA.  
 AC 09FG26;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE NON-LTR RETROELEMENT REVERSE TRANSCRIPTASE-LIKE.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 CC Eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-COLU-MBIA;  
 RA Kaneke T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF002032; BAB09815.1; -  
 KM RNA-directed DNA polymerase.  
 SQ SEQUENCE 676 AA; 76665 MW; ADIC1E1B846F53A CRC64;

Query Match 56.1%; Score 55.5; DB 10; Length 676;  
 Best Local Similarity 47.1%; Pred. No. 11;  
 Matches 8; Conservative 0; Mismatches 2; Indels 7; Gaps 1;

QY 1 RMP-----WMPKMP 10  
 |||||  
 Db 443 RSWPTLFALTVMWMPKMP 459

## RESULT 5

ID 09LON0 PRELIMINARY; PRT; 95 AA.  
 AC 09LON0;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE F5D14.5.  
 GN F5D14.5.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 CC Eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Liu S.X., Chan A., Yu G., Lee J.M., Lenz C., Pham P., Sakano H.,  
 RA Toriumi M., Vysotskaya V.S., Chin C., Chou J., Choi E., Chung M.,  
 RA Gonzalez A., Hwang B., Liu A., Vaysberg M., Altafi H., Brooks S.,  
 RA Buehler E., Chao Q., Conn L., Conway A.B., Hansen N.F.,  
 RA Johnson-Hopson C., Khan S., Kim C., Lam B., Miranda M.,  
 RA Palm C.J., Shinn P., Southwick A., Davis R.W., Ecker J.R.,  
 RA Federspiel N.A., Theologis A.;  
 RT "The sequence of BAC F5D14 from Arabidopsis thaliana chromosome 1.";  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Theologis A.;  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.

RC STRAIN-CV. COLUMBIA;  
RA Theologis A.;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. COLUMBIA;  
RA Theologis;  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AC007767; AAF81325.1;  
SQ SEQUENCE 95 AA; 10624 MW; 0C3985771E8B54E CRC64;

Query Match 53.5%; Score 53; DB 10; Length 95;  
Best Local Similarity 29.6%; Pred. No. 3.9;  
Matches 8; Conservative 2; Mismatches 3; Indels 14; Gaps 1;

OY 1 RRPWMP-----NRKPLI 13  
DB 59 RNNWMPVLYITDVGEMSWMMWMPVL 85

RESULT 6  
ID 080790 PRELIMINARY; PRT; 970 AA.  
AC 080790;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE REVERSE-TRANSCRIPTASE-LIKE PROTEIN.  
GN T12J2.5.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_Taxid=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. COLUMBIA;  
RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,  
RA Sykes S.M., Kaul S., Mason T.M., Kerlavage A.R., Adams M.D.,  
RA Somerville C.R., Venter J.C.;  
RT "Arabidopsis thaliana chromosome II BAC T12J2 genomic sequence."  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE  
TRANSCRIPTASE).  
CC EMBL: AC004483; AAC26674.1;  
DR InterPro: IPR000477; RVTse.  
DR Pfam: PF00078; rvt; 2.  
KW RNA-directed DNA polymerase.  
SQ SEQUENCE 970 AA; 110578 MW; C3A514F63BF67FA7 CRC64;

Query Match 53.0%; Score 52.5; DB 10; Length 970;  
Best Local Similarity 46.7%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 1; Indels 7; Gaps 1;

OY 3 WP-----WMPKW 10  
DB 739 WPTLFSGMIGIMWAKW 753

RESULT 7  
ID 082276 PRELIMINARY; PRT; 1231 AA.  
AC 082276;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE POTATIVE REVERSE TRANSCRIPTASE.  
GN T16B12.11.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_Taxid=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. COLUMBIA;  
RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,  
RA Shen M., Rensing C.M., Fraser C.M., Somerville C.R., Venter J.C.;  
RT "Arabidopsis thaliana chromosome II BAC T16B12 genomic sequence."  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE  
TRANSCRIPTASE).  
CC EMBL: AC005311; AAC63844.1;  
DR InterPro: IPR000477; RVTse.  
DR InterPro: IPR000169; Thiolprot\_act\_site.  
DR Pfam: PF00078; rvt; 1.  
DR PROSITE: PS00639; THIOI\_PROTEASE\_HIS; UNKNOWN.1.  
KW RNA-directed DNA polymerase.  
SQ SEQUENCE 1231 AA; 139476 MW; 5832FD60DF342B1 CRC64;

Query Match 53.0%; Score 52.5; DB 10; Length 1231;  
Best Local Similarity 46.7%; Pred. No. 45;  
Matches 7; Conservative 0; Mismatches 1; Indels 7; Gaps 1;

OY 3 WP-----WMPKW 10  
DB 1000 WPTLFSGMIGIMWAKW 1014

RESULT 8  
ID 09Y0E8 PRELIMINARY; PRT; 157 AA.  
AC 09Y0E8;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE SALIVARY GLAND SECRETION PROTEIN (FRAGMENT).  
GN SCS1 OR SCS-1 OR CG3047.  
GN SCS1 OR SCS-1 OR CG3047.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_Taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Roth G.E., Wattler S., Bornschein H., Lehmann M., Korge G.;  
RT "Structure and regulation of the salivary gland secretion protein gene  
Sgs-1 of Drosophila melanogaster."  
RL Genetics 0:0-0(1999).  
DR EMBL: AF156228; AAD43808.1;  
DR FlyBase: FBgn000372; Sgs1.  
DR InterPro: IPR002965; P-rich\_extensn.  
DR PRINTS: PRO1217; PRICHEXTENS.  
FT NON\_TER 1  
SQ SEQUENCE 157 AA; 17975 MW; 8495D86262819707 CRC64;

Query Match 52.5%; Score 52; DB 5; Length 157;  
Best Local Similarity 73.0%; Pred. No. 8.1;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 4 PWWMPKW 11  
DB 114 PQWPMSP 121

RESULT 9  
ID 0915M4 PRELIMINARY; PRT; 301 AA.  
AC 0915M4;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE HYPOTHETICAL PROTEIN PA0702.  
 GN PA0702.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 CC Pseudomonas.  
 OX NCBI\_Taxid=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.D., Brinkman F.S.L., Hultnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Gatter R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reiter J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 RT opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 DR EMBL; AE004506; AAC04091.1;  
 DR InterPro: IPR001541; Sterol\_desat.  
 KM Hypothetical protein: Complete proteome.  
 SQ SEQUENCE 301 AA; 33816 MW; 09FC2935E490ABD7 CRC64;

Query Match 52.5%; Score 52; DB 2; Length 301;  
 Best Local Similarity 70.0%; Pred. No. 15;  
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 WPMWPKMPL 12  
 11 11 11  
 DB 97 WPMWPKMPL 106

RESULT 10  
 Q097V5 PRELIMINARY; PRT; 1245 AA.  
 AC Q097V5;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE CONIDIOSPORE SURFACE PROTEIN.  
 GN CMPL.  
 OS Trichoderma harzianum.  
 OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Trichoderma.  
 OX NCBI\_Taxid=5544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 32173;  
 RA Puyesky M., Benhamou N., Ponce Noyola P., Bauw G., Ziv T.,  
 RA van Montagu M., Herrera Estrella A., Horwitz B.A.;  
 RT "Developmental regulation of a gene encoding a multidomain  
 RT conidiospore surface protein of Trichoderma, cmpl.";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ13651; CAB40845.1;  
 DR EMBL; 1245 AA; 135824 MW; 3249C749AFA0CDF8 CRC64;  
 SQ SEQUENCE

Query Match 52.5%; Score 52; DB 3; Length 1245;  
 Best Local Similarity 62.5%; Pred. No. 53;  
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 WPMWPKMPL 10  
 11 11 11  
 DB 1199 WPMWPKMPL 1206

RESULT 11  
 Q40548 PRELIMINARY; PRT; 109 AA.  
 AC Q40548;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE PISTIL-SPECIFIC EXTENSIN-LIKE PROTEIN PRECURSOR (FRAGMENT).  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; easterids I; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_Taxid=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. PETITE HAVANA; TISSUE-PISTILL;  
 RX MEDLINE=93005740; PubMed=1392607;  
 RA Goldman S., Pezzotti M., Seurinck J., Mariani C.;  
 RT "Developmental expression of tobacco pistil-specific genes encoding  
 RT novel extensin-like proteins.";  
 RL Plant Cell 4:1041-1051(1992).  
 CC -1- TISSUE SPECIFICITY: PISTIL (STIGMA AND STYLE TISSUE).  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSION BEGINS IN FLORAL BUDS AFTER PISTIL  
 CC DIFFERENTIATION AND LEVELS GRADUALLY INCREASE DURING FLOWER  
 CC DEVELOPMENT TOWARD ANTHESIS. LEVELS GRADUALLY DECREASE AFTER  
 CC POLLINATION AND ARE ABSENT BY THE SIXTH DAY AFTER POLLINATION.  
 DR EMBL; Z14014; CA78392.1;  
 DR Mendel; 16906; Nicita; 2747; 16906.  
 DR InterPro: IPR002965; P-rich\_extensin.  
 DR PRINTS; PR01217; PRICEXTENS.  
 KW Structural protein; Repeat; Glycoprotein; Signal.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 23 109 PISTIL-SPECIFIC EXTENSIN-LIKE PROTEIN.  
 FT DOMAIN 81 109 3 X 5 AA REPEATS OF S-P(4).  
 FT REPEAT 81 85 1.  
 FT REPEAT 93 97 2.  
 FT REPEAT 105 109 3.  
 FT NON\_TER 109  
 SQ SEQUENCE 109 AA; 12373 MW; 6A6BCFAC192AB8 CRC64;

Query Match 52.0%; Score 51.5; DB 10; Length 109;  
 Best Local Similarity 43.8%; Pred. No. 6.7;  
 Matches 7; Conservative 1; Mismatches 1; Indels 7; Gaps 1;

QY 3 WPMWPKMPL 11  
 11 11 11  
 DB 43 WPMWPKMPL 58

RESULT 12  
 Q008195 PRELIMINARY; PRT; 196 AA.  
 AC Q008195;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE CYSTEINE-RICH EXTENSIN-LIKE PROTEIN 2.  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; easterids I; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_Taxid=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=93342083; PubMed=8341705;  
 RA Wu H.M., Zou J., May B., Gu Q., Cheung A.Y.;  
 RT "A tobacco gene family for flower cell wall proteins with a proline-  
 RT rich domain and a cysteine-rich domain.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:6829-6833(1993).  
 DR EMBL; L13440; AAA34060.1;  
 DR Mendel; 16902; Nicita; 2747; 16902.  
 SQ SEQUENCE 196 AA; 21913 MW; 4C44E23C5B706E30 CRC64;

Query Match 52.0%; Score 51.5; DB 10; Length 196;  
 Best Local Similarity 43.8%; Pred. No. 11;  
 Matches 7; Conservative 1; Mismatches 1; Indels 7; Gaps 1;

OY 3 WPW-----WPMKWP 11  
 DB 40 WPMELPCVLTWMPFMP 55

RESULT 13  
 ID 008194 PRELIMINARY; PRT; 209 AA.

AC 008194; STRAIN-CV, COLUMBIA;  
 DT 01-JAN-1998 (TREMblrel. 05, Created)  
 DT 01-JUN-1998 (TREMblrel. 05, Last sequence update)  
 DE PISTIL-SPECIFIC EXTENSIN-LIKE PROTEIN PRECURSOR  
 DE (CYSTEINE-RICH EXTENSIN-LIKE PROTEIN 1).  
 OS Nicotiana tabacum (Common tobacco)  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 CC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
 NCBI\_TaxID=4097;

SEQUENCE FROM N.A.  
 RP MEDLINE=93342083; PubMed=8341705;  
 RA Wu H.M., Zou J., May B., Gu Q., Cheung A.Y.,  
 RT "A tobacco gene family for flower cell wall proteins with a proline-  
 RT rich domain and a cysteine-rich domain."  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:6829-6833(1993).  
 [2]

SEQUENCE OF 39-209 FROM N.A.  
 RP STRAIN-CV, PETITE HAVANA; TISSUE=PISTIL;  
 RX MEDLINE=93005740; PubMed=1392607;  
 RA Goldman S., Pezzotti M., Seurinck J., Mariani C.,  
 RT "Developmental expression of tobacco pistil-specific genes encoding  
 RT novel extensin-like proteins."  
 RL Plant Cell 4:1041-1051(1992).  
 CC -1- TISSUE SPECIFICITY: PISTIL (STIGMA AND STYLE TISSUE).  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSION BEGINS IN FLORAL BUDS AFTER PISTIL  
 CC DIFFERENTIATION AND LEVELS GRADUALLY INCREASE DURING FLOWER  
 CC DEVELOPMENT TOWARD ANTHESIS. LEVELS GRADUALLY DECREASE AFTER  
 CC POLLINATION AND ARE ABSENT BY THE SIXTH DAY AFTER POLLINATION.

DR EMBL, L13433; AAA34059.1; -  
 DR EMBL, Z14020; CAA78398.1; -  
 DR Mendel, 16901; Nicita;2747;16901.  
 KM Structural protein; Repeat; Glycoprotein; Signal.  
 FT STGNL 1 19  
 FT CHAIN 20 209  
 FT DOMAIN 81 126  
 FT REPEAT 81 85  
 FT REPEAT 93 97  
 FT REPEAT 105 109  
 FT REPEAT 117 121  
 FT REPEAT 122 126  
 FT CARBOHYD 146 146  
 SQ SEQUENCE 209 AA; 23351 MW; 12198BE2B8ED5 CRC64;

Query Match 52.0%; Score 51.5; DB 10; Length 209;  
 Best Local Similarity 43.8%; Pred. No. 12;  
 Matches 7; Conservative 1; Mismatches 1; Indels 7; Gaps 1;

OY 3 WPW-----WPMKWP 11  
 DB 45 WPMELPCVLTWMPFMP 60

RESULT 14  
 ID 09SJ28 PRELIMINARY; PRT; 321 AA.

AC 09SJ28; STRAIN-CV, COLUMBIA;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)  
 DE PUTATIVE NON-LTR RETROELEMENT REVERSE TRANSCRIPTASE.  
 DR AT3623250.

OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;

SEQUENCE FROM N.A.  
 RP STRAIN-CV, COLUMBIA;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,  
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,  
 RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RT "Sequence and analysis of chromosome II of Arabidopsis thaliana."  
 RL Nature 402:761-766(1999).  
 DR EMBL, AC006592; AAD2368.1; -  
 KW RNA-directed DNA polymerase.  
 SQ SEQUENCE 321 AA; 36192 MW; 43736712301D41BA CRC64;

Query Match 52.0%; Score 51.5; DB 10; Length 321;  
 Best Local Similarity 46.7%; Pred. No. 18;  
 Matches 7; Conservative 0; Mismatches 1; Indels 7; Gaps 1;

OY 3 WP-----WPMKWP 10  
 DB 90 WPMVFWAAMVWGMKW 104

RESULT 15  
 ID 056073 PRELIMINARY; PRT; 2970 AA.

AC 056073; STRAIN-CV, COLUMBIA;  
 DT 01-JUN-1998 (TREMblrel. 06, Created)  
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)  
 DE GENOME POLYPROTEIN.  
 OS Hepatitis GB virus A.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae.  
 NCBI\_TaxID=39112;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97126113; PubMed=8971037;  
 RA Leary T.P., Desai S.M., Yamaguchi J., Chalmers M.L., Schlauder G.G.,  
 RA Dawson G.J., Mushawar I.K.;  
 RT "Species-specific variants of GB virus A in captive monkeys."  
 RL J. Virol. 70:9028-9030(1996).  
 RN [2]

SEQUENCE FROM N.A.  
 RX MEDLINE=97437486; PubMed=9292019;  
 RA Leary T.P., Desai S.M., Erker J.C., Mushawar I.K.;  
 RT "The sequence and genomic organization of a GB virus A variant  
 RT isolated from captive tamarins."  
 RL J. Gen. Virol. 78:0-0(0).  
 RN [3]

SEQUENCE FROM N.A.  
 RX MEDLINE=98120818; PubMed=9460920;  
 RA Erker J.C., Desai S.M., Leary T.P., Chalmers M.L., Montes C.C.,  
 RA Mushawar I.K.;  
 RT "Genomic analysis of two GB virus A variants isolated from captive  
 RT monkeys."  
 RL J. Gen. Virol. 79:41-45(1998).  
 RN [4]

SEQUENCE FROM N.A.  
 RA Erker J.C., Desai S.M., Leary T.P., Chalmers M.L., Montes C.C.,  
 RA Mushawar I.K.;  
 RT Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO HEPATITIS C VIRUS RNA DEPENDENT RNA POLYMERASE.  
 DR EMBL, AF023424; AAC40501.1; -  
 DR HSP; P27958; 1HEI.

DR MEROPS; S29.002; .  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_NS5a.  
 DR InterPro; IPR000566; Lipocin\_cytfabp.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS3a; 1.  
 DR Pfam; PF00998; HCV\_RdRp; 1.  
 DR PROSITE; PS00213; LIPOCALIN; UNKNOWN\_1.  
 KW Nonstructural protein; Polypeptide; RNA-directed RNA polymerase.  
 SQ SEQUENCE 2970 AA; 322105 MW; 8B4F2FD9DB89F5A3 CRC64;

Query Match 52.0%; Score 51.5; DB 12; Length 2970;  
 Best Local Similarity 47.1%; Pred. No. 1.3e+02;  
 Matches 8; Conservative 0; Mismatches 2; Indels 7; Gaps 1;

QY 2 RWPWMP-----WKWP 11  
 ||||| | ||  
 Db 807 RWPWLPKMLVAAWMP 823

Search completed: January 30, 2002, 11:51:58  
 Job time: 197 sec